
**Pieter-Jan Maes, Marc Leman,
Katty Kochman, Micheline Lesaffre,
and Michiel Demey**

Institute for Psychoacoustics and
Electronic Music—Department of
Musicology
Ghent University
Blandijnberg 2
9000 Ghent, Belgium
{pieterjan.maes, marc.leman,
katty.kochman, micheline.lesaffre,
michiel.demey}@UGent.be

The “One-Person Choir”: A Multidisciplinary Approach to the Development of an Embodied Human–Computer Interface

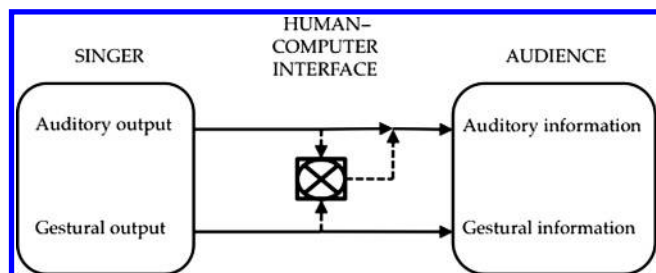
The “One-Person Choir” is a human–computer interface for singers that facilitates gestural control over a digital signal processing (DSP) module for harmonizing the singing voice in real time (see Figure 1). Harmonization adds extra pitch-shifted voices that are tonally related to the input voice. The interface captures global movements of the upper limbs by means of an integrated network of inertial sensors attached to the upper body of a singer. From these data, gestural cues are extracted and compared with a preconfigured gestural model that has been trained with empirical data. When the gestures of the singer match the preconfigured model, it is possible to control the harmonization of the singing input voice captured by a microphone. Thus, the interface allows a singer to naturally enhance the expressive qualities of his or her voice with the assistance of expressive gestures connected to an electronic environment.

The One-Person Choir can be integrated in interactive multimedia installations that exploit the expressive power of gestures in combination with singing. As will be argued in this article, installations illustrate, and elaborate on, an ongoing shift in contemporary electronic and electroacoustic music: the move from *interactive systems* (or hyperinstruments) to *composing interactions* (Di Scipio 2003).

Problem Definition

Research shows that gestures play an important role in the production and perception of singing performances. Gestures enhance the communication of intentionality and expressivity of singing (Davidson 2001; Yonezawa et al. 2006; Liao 2008; Luck and Toivainen 2008) in a way that is similar to the enhancement of speech (McNeill 2005). It is therefore not surprising that numerous human–computer interaction (HCI) applications have focused on facilitating gestural control over the quality of the singing voice (Hewitt and Stevenson 2003; Kessous 2004; Cook 2005; Yonezawa et al. 2005; D’Alessandro et al. 2006; Knapp and Cook 2006; D’Alessandro et al. 2007; Pérez, Knapp, and Alcorn 2007; Wong 2009). Gestural control over the voice has to be carried out in the electronic/digital domain, where microfeatures of expression can be exchanged between media (for example between gesture and audio). This could be done either by singing voice synthesis (Kessous 2004; Cook 2005; D’Alessandro et al. 2007; Wong 2009) or by transducing the acoustic voice into an electronic/digital signal by means of a microphone (Hewitt and Stevenson 2003; Yonezawa et al. 2005; Knapp and Cook 2006; Pérez, Knapp, and Alcorn 2007). Either way, once the voice is brought into the electronic/digital domain, the gesture that controls particular qualities of the voice (i.e., the mapping) can be arbitrarily chosen. A majority of the aforementioned interfaces establish the mediation between different modalities (basically from movement to sound) on a purely arbitrary basis (Kessous 2004; Cook 2005; Yonezawa et al. 2005; D’Alessandro et al. 2007; Wong 2009). However, this approach may impede the natural interaction

Figure 1. Structural overview of the One-Person Choir interface.



between the user and the digital sound “instrument” as well as between the performer and their audience. The few attempts that have been undertaken to create a natural gesture-to-sound mapping (Hewitt and Stevenson 2003; Knapp and Cook 2006; Pérez, Knapp, and Alcorn 2007) have had the disadvantage of not providing a firm empirical basis for the mapping.

Theoretical Framework

The One-Person Choir interface proposes a systematic, empirical solution to the gesture-to-sound mapping problem based on the embodied music cognition (EMC) paradigm (Leman 2007; Godøy and Leman 2010). This approach is grounded in multisensory integration, the coupling of perception and action, the study of motor imitation, and issues that relate to affect, emotions, and subjectivity (Camurri et al. 2005; Leman and Camurri 2006; Maes et al. 2010).

As stated earlier, a singing performance includes more than just auditory output: It explores and communicates ideas through vision and movement. According to the EMC theory it is through corporeal imitation that this multimodal stream of physical structures (e.g., audio, vision, movement) is translated into objects of a subjective *action-oriented ontology*, and vice versa (Leman 2007). The multisensory input received from the external physical world is associated with patterns of action. By internally mirroring these actions, they are experienced and understood as intentionally, expressively, and semantically meaningful. As such, this action-oriented ontology signals a kind of embodied, intermediate system that acts between (1) the purely

objective, physical reality and (2) the attribution of expressiveness and mental ideas to those physical signals. Because the experience and understanding of actions are partly shared by humans (based on biologically and cultural grounds), this ontology creates a repository for semantic communication (Leman 2007). As a result, the different types of sensory modalities involved in the production and perception of music should have a semantic match in relation to the action-oriented ontology of the musician and the audience. Globally stated, the main challenge is the search for an appropriate gestural mediation (and associated technology) that naturally expresses, and visually and kinesthetically communicates, the same idea as the musical output it produces. This multisensory congruency provides a natural feeling of causality in the human-computer interface.

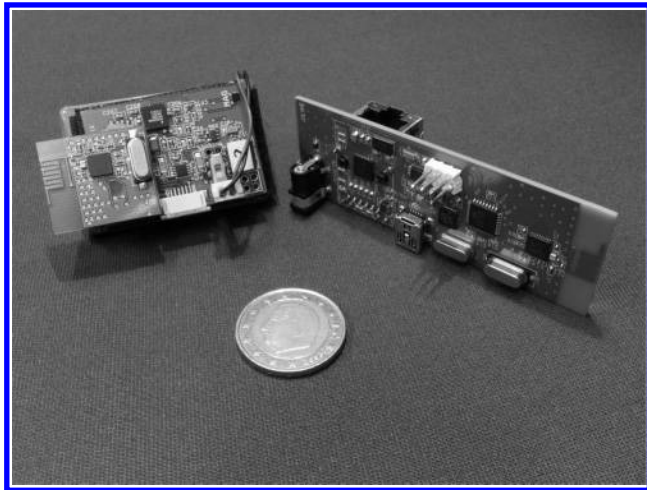
Methodology

The integration of the body and its expressive qualities in musical performance can only be achieved through a multi-layered analysis of objective and subjective performance components (Camurri et al. 2001). To that aim, this paper integrates low-level techniques adopted from computer and engineering disciplines. These techniques involve the measurement of physical body movements and, subsequently, the extraction of particular gestural features. In addition, perspectives on music interfaces are supported by empirical findings concerning high-level, subjective factors of musical involvement like action-perception coupling, intentionality, and meaning.

Bodily Motion Detection

Continuous movements of the upper limbs are the gestures that will be used to control voice harmonization. Therefore, we need to have a motion-detection system that delivers a low-level, physical representation of these movements. Concerning the motion-detecting system and the representation it delivers, a number of requirements need to be met.

Figure 2. HOP sensor (left) and receiver (right), with a coin (25.75-mm diameter) for comparison.



First, we need a three-dimensional (3-D) position description of the upper limbs. Second, this position must be expressed in reference to a coordinate system that is relative to the body itself. In other words, the position of the upper limbs needs to be expressed in reference to a person's *kinesphere* (peripersonal space)—that is, the space immediately surrounding the user's body and reachable by the limbs (Laban and Lawrence 1967; Farne, Dematte, and Ladavas 2005). Third, the motion-detection system may not impede spontaneous movement. Fourth, it must be easy to use the system in various performance contexts. Finally, the system must be invulnerable to varying light conditions, visual occlusion, and shadow issues. Because optical devices (video- or infrared-based) cannot adequately meet these requirements, inertial sensor technology seemed to be the best option.

Inertial Sensing Techniques: HOP Sensors

For the One-Person Choir interface we use state-of-the-art, custom-made wireless inertial motion detectors called HOP sensors (see Figure 2), which are named after the Hardware Ontwerp (i.e., Development) Project and produced by the Center for Microsystems Technology (CMST) at Ghent University (Kuyken et al. 2008; Huyghe, Doutreloigne, and Vanfleteren 2009). These sensors incorporate 3-D

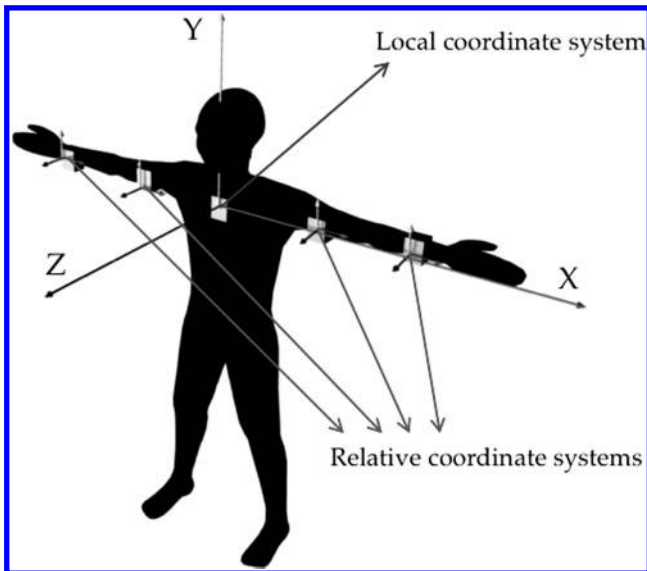
accelerometers combined with 3-D magnetometers. With the help of a wireless transceiver, the sensor is able to send the data from the accelerometers and magnetometers up to a range of 40 m at a sampling rate of 100 Hz via Ethernet to a computer. Due to the relatively small size of the sensor (55-mm long, 32-mm wide, and around 15-mm thick), the wireless transmission of data, and the standalone battery, sensors are easily attached with simple stretchable Velcro to parts of the singer's body, resulting in minimal restriction of bodily movement.

The preference for this sensor over similar, alternative ones (e.g., Shake SK7, Orient-2 [Young, Ling, and Arvind 2007], MTx XSens, Animazoo IGS-190) is motivated by an established collaboration of our research team with the CMST lab, where this type of sensor is used in ongoing research on inertial sensing technology. The implementation of this technology in the One-Person Choir functions as validation of the performance and usability of the technology. This close interaction between hardware creation and usability testing is indispensable for future development. Nonetheless, the One-Person Choir interface is created in such a way that it is an easy to use alternative, with similar inertial-sensor systems to sense movements of the upper limbs. This approach facilitates a broad implementation of the presented HCI application.

Position Estimation

The computational method that is used to calculate the 3-D position of the upper limbs in reference to the person's own peripersonal space is described in terms of *rigid body motion* and *forward kinematics*. A rigid body is considered as a system of particles, whereby the distances between all the particles of the system are fixed relative to each other. Therefore, we can conceive of the upper body as a kinematic chain of rigid bodies connected with joints characterized by a certain degree of freedom. For the One-Person Choir interface, it is sufficient to simplify this complex structure of the upper body to five rigid bodies: the torso, the two upper arms, and the two forearms. Now, by applying a forward kinematics algorithm—which is considered as the

Figure 3. 3-D visualization of a person equipped with the HOP sensors.



analytic geometry of motion of a biomechanical system of rigid bodies whereby position is estimated as a function of the joint angles—we can calculate, in real time, the 3-D position of the wrists and elbow joints as a function of the shoulder and elbow joint angles. Before the actual execution of the forward kinematics computation can be done, four conditions need to be met.

First, we have to obtain the 3-D orientation of each rigid body. To meet this condition, we attach five inertial sensors to the upper body in an integrated circuit, such that each rigid body is mounted with one sensor (see Figure 3). The 3-D accelerometer and 3-D magnetometer output of each sensor is then processed by a MATLAB implementation of an unscented Kalman filter (UKF), which is used to estimate the orientation of the sensor/rigid body. The pitch, yaw, and roll values (see Figure 4) that specify the orientation are shown in reference to an earth-fixed coordinate system.

Second, we have to obtain the relative differences in orientation between two succeeding rigid bodies. This is done by subtracting the corresponding pitch, yaw, and roll values of two sensors attached to the rigid bodies.

Third, the system needs to be calibrated. The intent is to establish a frame of reference that defines a specific posture in terms of the differences

in orientation found between the different sensors mounted to the rigid bodies. Therefore, a person using the One-Person Choir interface is asked to form a T-shape with the torso and the arms and is then asked to turn the inside of their hands forward (see Figure 3). In this posture, all sensors are correctly positioned so that the orientation of their coordinate system can be equalized.

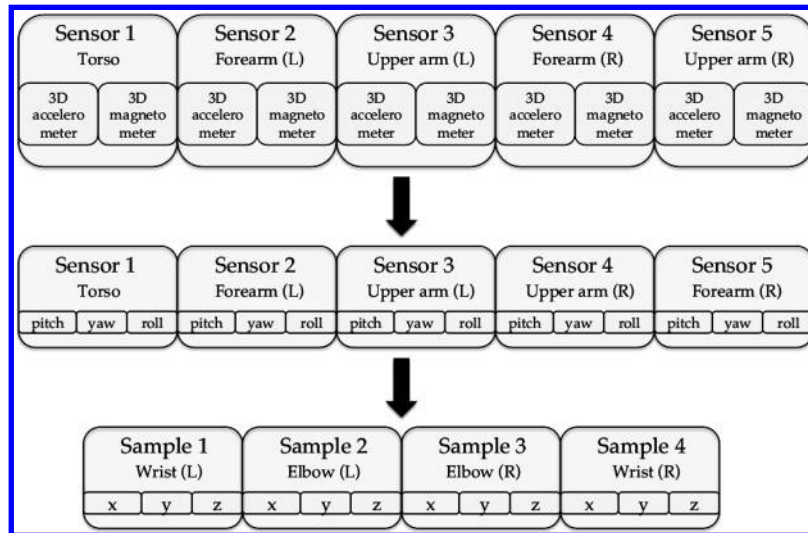
We have already mentioned that estimating the position of the upper limbs is only relevant in relation to the peripersonal sphere of the body itself. As such, we need to obtain a coordinate system that is relative to body rotation and displacement. We choose to define this local coordinate system, which moves along with the body, by placing its origin at the sensor attached to the torso. Figure 3 shows an example configuration of the x , y , and z axes relative to the person's body. The time-varying positions of the wrists and elbows will be determined with respect to this local-coordinate system.

Once the four conditions are met, the actual forward kinematics algorithm can be performed. This is done in real time on the Max/MSP platform. An algorithm that integrates trigonometric mathematics calculates the position of the upper limbs based on the relative changes in orientation between the relative-coordinate systems (on the upper limbs) and the local-coordinate system (on the torso). For a more detailed explanation of the internal functioning of the algorithm, see Maes et al. (2010).

Movement Processing—Feature Extraction

The proposed HCI provides gestural control over harmonizer-generated, pitch-shifted voices, each sounding above the original monophonic singing voice. (Throughout this article, the term “monophonic” refers to the number of simultaneous musical parts, not the number of audio channels.) To realize this, the raw movement data need further processing. In what follows, two features are considered relevant as input to the sound interface. The first is the amount of contraction and expansion, within a person's peripersonal space, of the upper limbs; the second is the direction of this movement

Figure 4. Real-time processing stream of data (100 Hz) produced subsequently by: the inertial sensors (top), the UKF (middle), and the Max/MSP forward kinematics algorithm (bottom).



feature in the peripersonal space. The choice for these particular movement features is motivated by findings of previous research indicating that (1) the dynamic nature of movement is an important aspect in affective communication (Jellema and Perrett 2006), (2) the upper-body features are most significant in conveying emotion and expressiveness (Kleinsmith, Fushimi, and Bianchi-Berthouze 2005), (3) the movement size and openness can be related to the emotional intensity of the musical sound production (Davidson 1994; Camurri et al. 2004), and (4) an open body position, in contrast to a closed body position, reinforces the communicator's intent to persuade (Mehrabian and Friar 1969; McGinley, LeFevre, and McGinley 1975).

First Movement Feature: Contraction Index

The spatio-kinetic movement cue represented by the contraction index of the upper limbs gives a measure of the amount of the peripersonal space that is used by a person. Similar to the contraction index defined by Camurri, Lagerlöf, and Volpe (2003), the contraction index we use results in a value between zero and one. Unlike the algorithm used by Camurri, Lagerlöf, and Volpe, the contraction index here is defined by two separate measurements. From the positional data of the upper limbs (see Figure 4,

bottom), the Euclidean distance is calculated between (1) the elbows relative to each other and (2) the wrists relative to each other. The resulting values are then normalized between zero and one. As a result, we obtain a more nuanced estimate of the amount of space used by a person.

Second Movement Feature: Direction of Movement

The second movement feature that will be used to control the One-Person Choir is the direction in which the expansion and contraction of the upper limbs is performed. While creating the computational method to extract this feature, we adopted the idea, after Laban (1963), that directions of movement radiate from the center of one's peripersonal space and, as such, have to be determined in relation to one's own body. The method is subdivided into an off-line and on-line process.

Off-line Operation

During the off-line process, a computational representation of a person's kinesphere is created. It consists of a sphere with a radius of unit length in which the center and x , y , and z dimensions correspondingly coincide with the origin and x , y , and z axes of the local-coordinate system (see Figure 3).

Figure 5. Representations of (left) the azimuth plane, (center) the colatitude plane, and (right) the 2-D representation of the kinesphere as a function of the spherical coordinates.

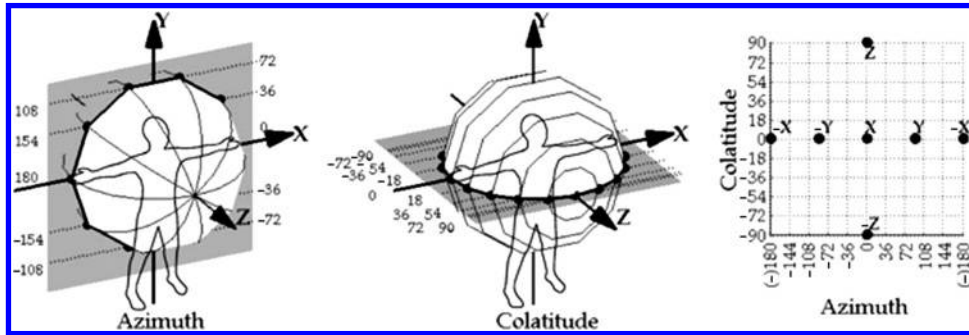
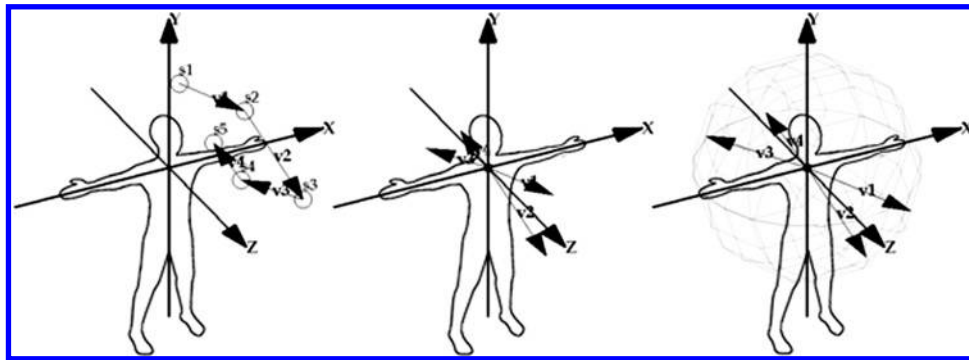


Figure 6. On-line process for obtaining directional vectors: (left) Five successive samples ($s1 - s5$), between which directional vectors

($v1 - v4$) are drawn, (center) directional vectors shifted to the origin, and (right) directional unit vectors.



In this kinesphere, every direction radiating from the center is represented by a vector that starts from the origin and extends to a point on the surface of the sphere. Every end point can then be defined in terms of a spherical coordinate, consisting of an azimuth value Θ (theta) and colatitude value Φ (phi) value. The vertical $x - y$ plane forms the azimuth plane, and the horizontal $x - z$ plane forms the colatitude plane (see Figure 5). The azimuth and colatitude values are both measured with respect to the positive X -axis.

The surface of the sphere is further subdivided into a manageable amount of *directional segments*. The number of segments (i.e., the resolution of the sphere) is by default 10×10 . Each of the 100 segments is then numbered and defined in terms of a unique pair of spherical coordinates, defining its maximum and minimum values of azimuth and colatitude.

On-line Operation

The on-line operation facilitates the real-time calculation of the direction of movement for each part of the upper limbs (wrists, elbows, etc.). The computational method that performs this operation is theoretically founded on Rudolf Laban's concepts. As mentioned before, Laban (1963) assumes that in relation to our body we have the feeling that directions radiate from the center of our kinesphere. In line with this idea, we will represent the direction of movement as a unit vector that starts at the origin of the local-coordinate system (i.e., the center of the kinesphere) and extends to a point on the surface of the kinesphere (see the previous section Off-line Operation). A threefold process leads to this representation (see Figure 6). First, for each incoming 3-D position sample acquired by the inertial sensor system, a vector is drawn from the previous sample to the new one. Second, the obtained vector is

shifted to the origin of the local-coordinate system. Third, the magnitude of the vector bounded to the origin is normalized to a unit magnitude. Finally, from the Cartesian representation of this directional unit vector, we calculate the spherical coordinate in terms of azimuth and elevation values. Additionally, we check what segment of the kinesphere is intersected by the directional vector. This operation is repeated at the same rate as the 3-D position estimation rate (i.e., 100 Hz). Because the first directional vector can be calculated only after the input of the second position sample, there is an initial but negligible delay of one sample.

Gesture-to-Sound Mapping

After the development of real-time methods for extracting specific gestural cues—namely, the contraction index and the direction of movement—the next challenge we face is to propose a mapping model that links the gestural cues to specific musical cues. As specified in the introduction, the purely technological aspects of the interface design (e.g., bodily motion detection, feature extraction) need to take into consideration subjective phenomena like multisensory integration, the coupling of perception and action, the study of motor imitation, and issues that relate to affect and emotions. By elaborating on previous empirical research (Maes et al. 2010), we will propose what we believe is the appropriate gestural model for controlling the harmonization of the singing voice.

Experimental Approach to the Gesture-to-Sound Relationship

This section relies on measurements and results provided by the experimental research of Maes et al. (2010). In the experiment, subjects ($n = 25$) were asked to listen to four pre-recorded sound stimuli and corporeally imitate the perceived characteristics. That study is especially relevant to the One-Person Choir interface because of the musical nature of the stimuli. The stimuli focused on the musical effect generated by a harmonizer—that is,

the gradual addition and disappearance of extra, pitch-shifted voices related to an originally monophonic input voice. While corporeally imitating this musical structure, it seemed that the subjects shared common gestural patterns. The crucial idea behind the gesture-to-sound mapping of the One-Person Choir is that the integration of these gestural patterns provides a natural and intuitive means to effect harmonization of the singing voice.

Analysis of the Contraction Index

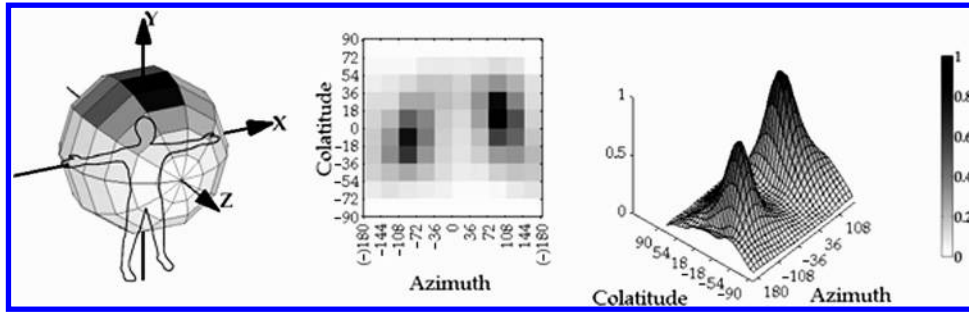
Statistical analysis of the collected movement data showed that the addition (or removal) of extra, harmonic voices tended to be corporeally imitated by subjects by expanding (or, respectively, contracting) their upper limbs within peripersonal space, and tended to be perceived as having a higher (or, respectively, lower) emotional intensity. The analysis was made with the method outlined in the section First Movement Feature: Contraction Index. We observed that the expansion is primarily due to an expanding distance between the elbows (i.e., outward movement of the upper arm). The influence of the wrists on the expansion of the upper limbs in the peripersonal space was not significant. Therefore, the One-Person Choir only includes the contraction index defined by the distance between the elbows.

Analysis of the Direction of the Expansion/Contraction Movements

To prepare for the actual analysis, a method was developed to determine which directions are most frequently used in a pre-recorded movement trajectory. This is established by creating a *direction density matrix* (DDM) of which the dimensions correspond with the resolution of the computational representation of the kinesphere (see the section Second Movement Feature: Direction of Movement).

For a recorded movement trajectory of n samples in the format specified in Figure 4, $n - 1$ directional vectors can be calculated, each crossing one segment of the kinesphere. This allows creating a DDM in which each element corresponds to a specific segment of the kinesphere. Initially, zero is assigned to

Figure 7. 3-D and 2-D visualizations of a direction density matrix (DDM).



each element. But each time a segment is intersected by a directional vector, one is added to the value of the element assigned to that particular segment. Finally, after the last directional vector of the movement trajectory is processed, the value assigned to each element of the DDM is equal to the number of times the corresponding segment has been intersected. The values are then normalized between zero and one. These normalized values are called *directional indices* (DIs). In Figure 7, three-dimensional and two-dimensional (2-D) visualizations of a DDM and corresponding DIs are shown.

This method was used in the analysis of the data obtained during the experiment of Maes et al. (2010), to see if there were commonalities among subjects regarding the direction in which the expansion of movement took place. As stated previously, we take into account the direction of movement of both elbows. Basically, this means that we have to do the same analysis twice. For simplicity, we will only discuss the analysis of the left elbow. Afterwards, the results of both elbows will be discussed.

Because each subject ($n = 25$) moved in response to four different sound stimuli, there were 100 different movement trajectories performed by the left elbow that will be taken into account. For each of the 100 movement trajectories, a DDM was created. The mean DDM across all the performances was then calculated. This DDM showed a clear concentration of activity, specified by an azimuth in degrees between 108° and 144° and a colatitude between 18° and 36° . These results suggest a strong commonality regarding the directionality of the expansion/contraction motor response of subjects. The same analysis process was executed for the right

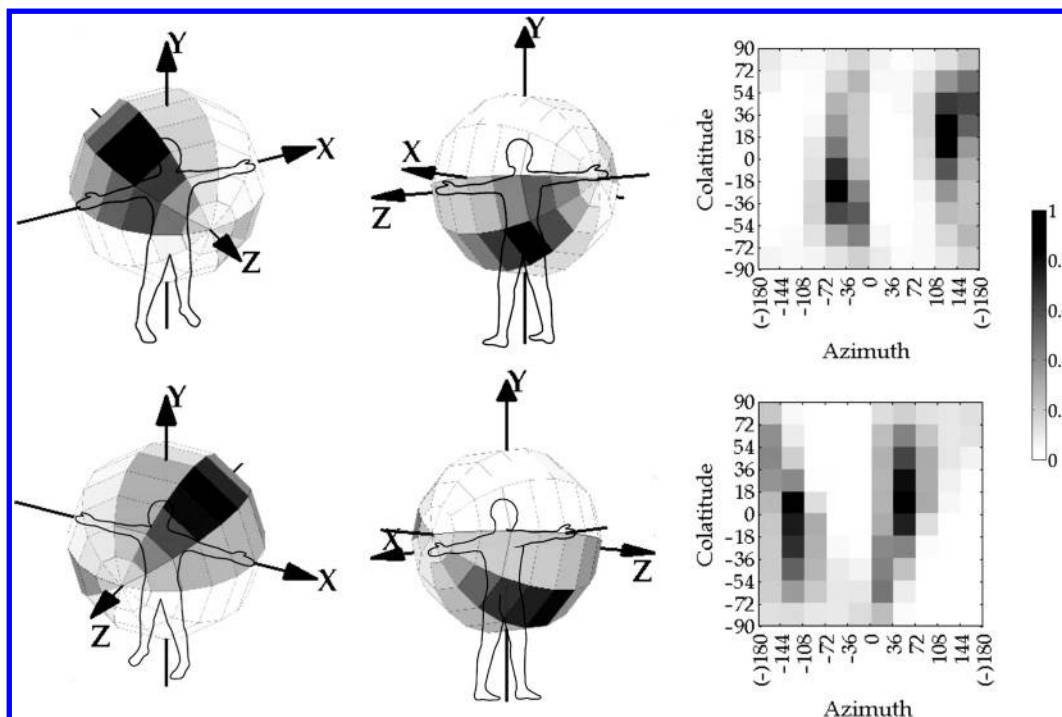
elbow leading to similar results and interpretations (see Figure 8).

Gestural Model

In relation to the development of the One-Person Choir application, there are two conclusions that can be drawn from these empirical results. First, the sound-synthesis process of adding voices to a monophonic input is spontaneously imitated by corporeal activity, characterized by the expanded movement of the upper arms. Second, the majority of participants shared the same direction of expansion. From these results, a gestural model is trained and integrated in the computational algorithm implemented in the Max/MSP environment. The specific properties of the model are internally mapped to the different parameters of the sound-synthesis module. As a result, when singers move according to the model, it is possible for them to alter their own voice in correspondence to the mapping configuration.

The interface system was developed to calculate in real time both the varying distance between the two elbows, specified as the *contraction index* (CI), and the direction in which an increase in distance between the elbows (i.e., expansion of the upper arms) takes place. The distance between the elbows when the arms are hanging loosely against the singer's sides corresponds to a CI of 1, the maximum distance between the elbows to a CI of 0. The volume of two extra voices, created in real time by the harmonizer and mixed with the captured voice of the singer, is regulated by the values of the CI. If

Figure 8. 3-D and 2-D visualization of the mean DDMs (top: right elbow; bottom: left elbow), indicating the directional index (DI) per directional segment.



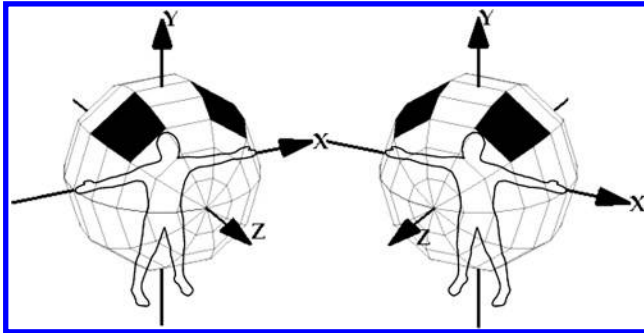
the CI equals 1, then there are no extra voices. If the CI equals 0.5, then the maximum volume of the first extra voice is reached, while there is no extra second voice. If the CI equals 0, then the two extra voices both reach their maximum volume. However, an increase in distance between the elbows results in volume changes of the extra, harmonized voices only when the elbows move toward the spherical coordinates specified in the gestural model. These spherical coordinates are chosen in correspondence with the results of the aforementioned experimental study (see the section Off-line Operation). However, the regions of maximal density that cover only a small portion of the directional sphere are enlarged to permit variation in the singer's movement (see Figure 9). The azimuth value of the left elbow is fixed between 36° and 72° and the colatitude value between 0° and 36° . The azimuth value of the right elbow is fixed between 108° and 144° and the colatitude value between 0° and 36° .

The direction feature thus acts as a switch that closes (i.e., admits a connection) when the input corresponds to the model and opens (i.e., prohibits

a connection) when there is no correspondence (see Figure 10). On the other hand, a decrease in distance between the elbows results in volume changes of the extra, harmonized voices regardless of the direction of movement. This is done to increase the user's freedom and to avoid jumps in the control signal.

In this case, the gestural model was trained with movement data obtained by experimental research to enhance the usability and intuitiveness of the application. As such, the model describes a simple but highly effective gesture to harmonize a singer's voice and subsequently to increase musical expressiveness. However, with only small and basic adjustments, it is possible for the user to deviate from the preconfigured model and adjust it to his or her personal needs. Moreover, the system facilitates the integration of other gestural models capable of controlling additional sound parameters. Another advantage of the HCI application is the possibility to use movement-sensing and -capturing systems other than the HOP sensors used in this study. The only requirement is that the systems must be able to output the (relative) 3-D position of specific points

Figure 9. 3-D visualization of the gestural model. The harmonizer is activated when the system detects elbow movement toward the directional segments depicted by dark regions.



of the human body in real time. This makes the One-Person Choir a dynamic, flexible, and user-centered HCI application that can easily be integrated in a music performance context.

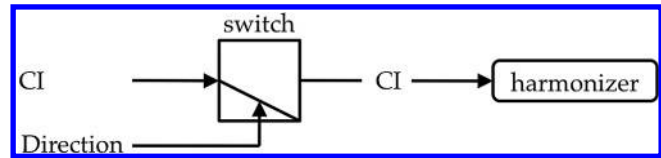
Interaction Designs

To test the embodied interface, different interaction designs were tried out in concrete artistic performances. In what follows, we give a brief overview of these designs.

Design One

The first interaction design (see Figure 11a) consisted of a solo singer equipped with the One-Person Choir while performing some existing vocal pieces (*Summertime* by George Gershwin, *Ave Maria* by Johann Sebastian Bach, etc.). The singer's gestures were sensed by the inertial sensor device and mapped to the harmonizer DSP, as explained in the earlier section Gestural Model. The harmonizer detected the singer's pitch and added the intervals of a third and a fifth in correspondence with a specified musical key. Specific configurations could also be made for pieces without specific tonalities. For this design, the singer could select the key in real time via an in-house created, non-commercial data glove. A micro-switch is located at the end of each finger with which a user can select a pre-configured key via the Musical Instrument Digital Interface (MIDI) protocol. This allowed for harmonic modulation within the musical piece. This design functioned as

Figure 10. Visualization of the interaction between the two properties (i.e., CI and direction) defined in the gestural model.



a practical test bed for evaluating the usability of the One-Person Choir interface and the integrated mapping strategy. The singer praised the gesture-to-sound mapping for its ability to be used in a very natural way and for creating a sense of augmented awareness of multisensory HCI based on expressive gesture.

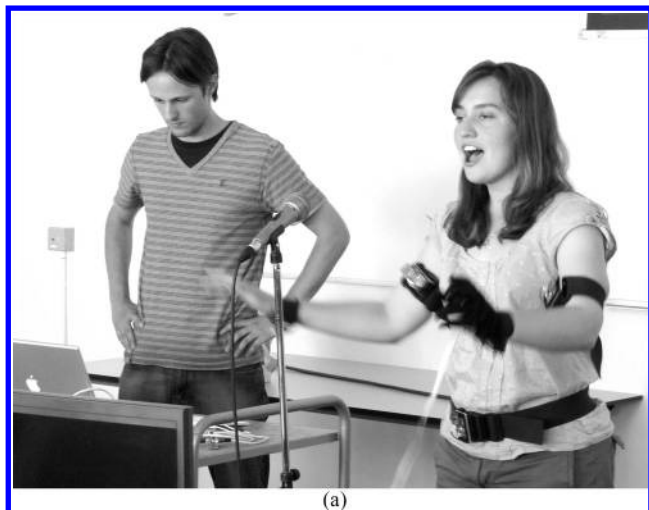
Design Two

The second interaction design (see Figure 11b) was very similar to the previous design, but different in that the singer was accompanied by other musicians. Also, the ensemble performed a different piece of music: the arrangement of the folk song *Black is the Color of My True Love's Hair* by Luciano Berio. A few minor adjustments were made to the previous design on the DSP level. For example, we implemented an additional parameter in the mapping structure and the harmonization of the voice was combined with more reverb. This gave the voices a warm and full overall sound that mixed well with the instrument sounds. Finally, because the data glove was wired and, as such, constrained free movement within the performance space, an external person selected the musical key, using the harmonizer software in real time.

Design Three

The third interaction design resulted from a commission for a new composition using the One-Person Choir interface. The composition was written by Olmo Cornelis for two soprano voices, dancer, and electronics, and was entitled *Nelumbo*. The concept of this design is radically different from the previous two. The motion-sensing system is not worn by the singers but by the dancer (see Figure 12). The dancer, using the gesture-to-sound mapping, controls with

Figure 11. The One-Person Choir implemented in a performance context: (a) Design One and (b) Design Two.



her right hand the volume of two extra voices added to the singer on her right; similarly, her left hand controls the volumes of two extra voices added to the singer on her left. When the distance between the hand and the chest equals zero, there are no extra voices added. When the distance between the hand and the chest is greatest, the three voices reach full volume. When the hand is placed in the middle, only the first two voices reach maximum volume while the third is silent. The added voices are pitch-shifted, tonally related duplications of the singer's voice; the second voice adds a third while the third voice adds a fifth. The composition was written in such a way that the additions of the extra voices were harmonically interesting. Some passages were intentionally tonal and mixed well with the added voices, while other passages were

Figure 12. 14 March 2010: (composer), Pieter-Jan Maes (electronics), Chia-Fen Wu (singer), Antonella Cusimano (dancer), and Katty Kochman (singer).



especially assembled to create tension in the music. Using the harmonizer software, the composer performed the key selection in real time.

This approach expanded the interaction possibilities of the One-Person Choir interface in a profound way (see the Discussion). Moreover, it radically alters the traditional view of the dancer as subordinate to the music. Now, the dancer is empowered with direct control over the auditory result. The expressive power of spontaneous dance gestures is employed to control the expressive content of the music. This enhances not only the communication of emotions to the audience but also boosts the interaction and collaboration between the dancers and singers. This idea of empowering the dancer to manipulate musical processes was previously explored by Siegel and Jacobsen (1998).

Discussion

In this section, we discuss the position and contribution of the One-Person Choir interface, as well as the presented interaction designs, in the context of electroacoustic and computer music history.

Interface Design

From the 1970s through the 1990s, a shift from procedural, algorithmic-music systems to real-time interactive systems (Wegner 1997; Goldin, Smolka, and Wegner 2006) occurred, enabling user-based control over the algorithmic parameters regulating

the sonic output. Soon, it became apparent that motor and perception issues must be taken into account to enhance the usability of HCI designs (Grudin 1990; Vaggione 2001; Beaudouin-Lafon 2004). In order for an HCI to facilitate natural control, the action performed by the user and the perception of the sonic output effectuated by this action must be tightly coupled. However, in practice, HCI designs often integrate a gesture-to-sound mapping that is based purely on arbitrary decisions, constraining natural, intuitive control. The design of the One-Person Choir interface offers an original solution to this mapping problem by adopting an embodied approach to music production and perception. Based on empirical findings (Maes et al. 2010), the visual, auditory, and kinesthetic modalities are made congruent, thus enabling a natural feel of causality in the human-computer interface.

Interaction Design: From Interactive Systems to Composing Interactions

The first two interaction designs belong to a now-standard repertoire of interactive music systems according to which—in the words of Di Scipio (2003)—a human agent selects and activates particular functions and processes whose output sample streams are linearly summed together.

Composer/researchers like George E. Lewis and Agostino Di Scipio contributed to an interesting evolution in the conceptualization of interaction designs (Di Scipio 2003, 2005; Meric and Solomos 2009). Particularly interesting was what Di Scipio (2003, p. 271) calls “a shift from creating wanted sounds via interactive means towards creating wanted interactions having audible traces.” According to Agostino Di Scipio’s interaction model, the very process of interaction is not a matter of a linear communication flow from an agent to some computer algorithms that it controls—very much like our first two interaction designs—but is merely “a by-product of lower-level interdependencies among system components” (Di Scipio 2003, p. 271). Our third interaction design extensively contributes to this shift from interactive music composing towards

composing musical interactions. Whereas Di Scipio (2003) places the interaction merely at the sonic signal level, we envision the core of the interdependency between components at the motor level. This provides a basis for a multimodal exchange of information.

The primary motor components for this third interaction design consist of the two singers’ vocal apparatuses and the dancer’s body, extended with the HCI. They are conceived as active mediators translating mental phenomena (e.g., intentions, ideas, feelings, moods) into encodings of multimodal physical energy. As such, the artistic result—which comprises auditory structures (original voices and augmented voice) as well as visual, kinesthetic, and possibly tactile structures—can be seen as an epiphenomenon, emerging from a trajectory of constrained embodied interactions among the different components of the system and the composer’s compositional structure.

Whereas the EMC theory focuses on meaningful articulations (i.e., gestures) of the individual body, the focus on the integrated whole of mutually influencing interpersonal bodily articulations extends the communicative qualities of the human body into a complex “social body.” Because it is a well-known fact in social psychology that movements play an essential role in social information processing (Barsalou et al. 2003; Morganti 2008), we can say that what is actually generated by the interaction design is the shaping and sharing of musical thoughts and feelings in a social and collaborative context, one enhanced by HCI technology. The dialogue between the different components could not be anticipated in a straightforward, linear way because it is heavily dependent on factors like personality, mood, social context, physical environment, and so on. As such, the artistic output emerges much more from the embodied experience of an augmented social collaboration than as a fixed product of a purely cognitive effort.

Conclusion

Our human-computer interface design contributes in multiple ways to ongoing artistic praxis and academic research. First, the One-Person Choir

presents an embodied human-computer interface for music that incorporates an original solution for the traditional mapping problem in electronic and digital human-computer interfaces. Second, we extended the interaction model of Di Scipio (2003) by focusing on the motor level, the multimodal character of music, and the social interaction. Third, our notion of the social body extends the present EMC theory which focuses almost exclusively on the individual body. However, further empirical research must be conducted to make more fundamental contributions.

Acknowledgments

This work was carried out in the context of the EmcoMetecca project (Leman 2008) supported by the Flemish government. The authors want to thank Raven van Noorden for her contribution to the 3-D visualizations.

References

- Barsalou, L. W., et al. 2003. "Social Embodiment." In B. H. Ross, ed. *The Psychology of Learning and Motivation*. San Diego, California: Academic Press, pp. 43–92.
- Beaudouin-Lafon, M. 2004. "Designing Interaction, Not Interfaces." In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pp. 15–22.
- Camurri, A., et al. 2001. "A Multi-Layered Conceptual Framework for Expressive Gesture Applications" In *Proceedings of the MOSART Workshop on Current Research Directions in Computer Music*, pp. 29–34.
- Camurri, A., et al. 2004. "Multimodal Analysis of Expressive Gesture in Music and Dance Performances." In A. Camurri and G. Volpe, eds. *Gesture-Based Communication in Human-Computer Interaction*. Berlin: Springer Verlag, pp. 20–39.
- Camurri, A., et al. 2005. "Communicating Expressiveness and Affect in Multimodal Interactive Systems." *IEEE Multimedia*, 12(1):43–53.
- Camurri, A., I. Lagerlöf, and G. Volpe. 2003. "Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques." *International Journal of Human-Computer Studies*, 59(1–2):213–225.
- Cook, P. R. 2005. "Real-Time Performance Controllers for Synthesized Singing." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 236–237.
- D'Alessandro, N., et al. 2006. "Real-Time Calm Synthesizer: New Approaches in Hands-Controlled Voice Synthesis." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 266–271.
- D'Alessandro, N., et al. 2007. "Realtime and Accurate Musical Control of Expression in Singing Synthesis." *Journal on Multimodal User Interfaces* 1(1):31–39.
- Davidson, J. W. 1994. "What Type of Information Is Conveyed in the Body Movements of Solo Musician Performers." *Journal of Human Movement Studies* 26(6):279–301.
- Davidson, J. W. 2001. "The Role of the Body in the Production and Perception of Solo Vocal Performance: A Case Study of Annie Lennox." *Musicae Scientiae* 5(2):235–256.
- Di Scipio, A. 2003. "'Sound Is the Interface': From Interactive to Ecosystemic Signal Processing." *Organised Sound* 8(3):269–277.
- Di Scipio, A. 2005. "Due Di Uno: A Composition Dedicated to Horacio Vaggione." *Contemporary Music Review* 24(4):383–397.
- Farne, A., M. L. Dematte, and E. Ladavas. 2005. "Neuropsychological Evidence of Modular Organization of the near Peripersonal Space." *Neurology* 65(11):1754–1758.
- Godøy, R. I., and M. Leman. 2010. *Musical Gestures: Sound, Movement, and Meaning*. New York: Routledge.
- Goldin, D., S. Smolka, and P. Wegner, eds. 2006. *Interactive Computation: The New Paradigm*. London: Springer.
- Grudin, J. 1990. "The Computer Reaches Out: The Historical Continuity of Interface Design." In *Proceedings of the SIGCHI 1990 Human Factors in Computing Systems Conference*, pp. 261–268.
- Hewitt, D., and I. Stevenson. 2003. "E-Mic: Extended Mic-Stand Interface Controller." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 122–128.
- Huyghe, B., J. Doutrelouigne, and J. Vanfleteren. 2009. "Design of Flexible, Low-Power and Wireless Sensor Nodes for Human Posture Tracking Aiding Epileptic Seizure Detection." In *Proceedings of the Annual IEEE Conference on Sensors*, pp. 1963–1966.
- Jellema, T., and D. I. Perrett. 2006. "Neural Representations of Perceived Bodily Actions Using a Categorical Frame of Reference." *Neuropsychologia* 44(9):1535–1546.

- Kessous, L. 2004. "Gestural Control of Singing Voice, a Musical Instrument." In *Proceedings of the Sound and Music Computing Conference*. Available on-line at smcnetwork.org/node/1040. Accessed January 2011.
- Kleinsmith, A., T. Fushimi, and N. Bianchi-Berthouze. 2005. "An Incremental and Interactive Affective Posture Recognition System." In *International Workshop on Adapting the Interaction Style to Affective Factors*. Available on-line at www.di.uniba.it/intint/UM05/list-ws-um05.html. Accessed January 2011.
- Knapp, R. B., and P. R. Cook. 2006. "Creating a Network of Integral Music Controllers." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 124–128.
- Kuyken, B., et al. 2008. "The Hop Sensor: Wireless Motion Sensor." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 229–231.
- Laban, R. 1963. *Modern Educational Dance*. London: MacDonald and Evans.
- Laban, R., and F. C. Lawrence. 1967. *Effort*. London: Macdonald and Evans.
- Leman, M. 2007. *Embodied Music Cognition and Mediation Technology*. Cambridge, Massachusetts: MIT Press.
- Leman, M. 2008. "Emcometecca Project." Available on-line at www.ipem.ugent.be/?q=EmcoMetecca. Accessed January 2011.
- Leman, M., and A. Camurri. 2006. "Understanding Musical Expressiveness Using Interactive Multimedia Platforms." *Musicae Scientiae* 10(1):209–234.
- Liao, M. Y. 2008. "The Effects of Gesture Use on Young Children's Pitch Accuracy for Singing Tonal Patterns." *International Journal of Music Education* 26(3):197–211.
- Luck, G., and P. Toiviainen. 2008. "Exploring Relationships between the Kinematics of a Singer's Body Movement and the Quality of Their Voice." *Journal of Interdisciplinary Music Studies* 2(1–2):173–186.
- Maes, P.-J., et al. 2010. "From Expressive Gesture to Sound: The Development of an Embodied Mapping Trajectory inside a Musical Interface." *Journal on Multimodal User Interfaces* 3(1–2):67–78.
- McGinley, H., R. LeFevre, and P. McGinley. 1975. "The Influence of a Communicator's Body Position on Opinion Change in Others." *Journal of Personality and Social Psychology* 31(4):686–690.
- McNeill, D. 2005. *Gesture and Thought*. Chicago: University of Chicago Press.
- Mehrabian, A., and J. T. Friar. 1969. "Encoding of Attitude by a Seated Communicator Via Posture and Position Cues." *Journal of Consulting and Clinical Psychology* 33(3):330–336.
- Meric, R., and M. Solomos. 2009. "Audible Ecosystems and Emergent Sound Structures in Di Scipio's Music. Music Philosophy Helps Musical Analysis." *Journal of Interdisciplinary Music Studies* 3(1–2):57–76.
- Morganti, F. 2008. "What Intersubjectivity Affords: Paving the Way for a Dialogue between Cognitive Science, Social Cognition and Neuroscience." In F. Morganti, A. Carassa, and G. Riva, eds. *Enacting Intersubjectivity: A Cognitive and Social Perspective on the Study of Interactions*. Amsterdam: IOS Press, pp. 3–16.
- Pérez, M. A. O., B. Knapp, and M. Alcorn. 2007. "Díamair: Composing for Choir and Integral Music Controller." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 289–292.
- Siegel, W., and J. Jacobsen. 1998. "The Challenges of Interactive Dance: An Overview and Case Study." *Computer Music Journal* 22(4):29–43.
- Vaggione, H. 2001. "Some Ontological Remarks About Music Composition Processes." *Computer Music Journal* 25(1):54–61.
- Wegner, P. 1997. "Why Interaction Is More Powerful Than Algorithms." *Communications of the ACM* 40(5):80–91.
- Wong, E. L. 2009. "Augmenting Media Performance with Interactive Technology." *Proceedings of the Generative Art Conference*, pp. 283–293.
- Yonezawa, T., et al. 2005. "Handysinger: Expressive Singing Voice Morphing Using Personified Hand-Puppet Interface." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 121–126.
- Yonezawa, T., et al. 2006. "Cross-Modal Coordination of Expressive Strength between Voice and Gesture for Personified Media." In *Proceedings of the International Conference on Multimodal Interaction*, pp. 43–50.
- Young, A. D., M. J. Ling, and D. K. Arvind. 2007. "Orient-2: A Realtime Wireless Posture Tracking System Using Local Orientation Estimation." In *Proceedings of EmNets*, pp. 53–57.